

**Georgia State University**  
**ScholarWorks @ Georgia State University**

---

Philosophy Theses

Department of Philosophy

---

8-7-2018

# Isolated Sacred Value Theory: An Account of Moral Conative Attitudes

Adrian Pecotic

Follow this and additional works at: [https://scholarworks.gsu.edu/philosophy\\_theses](https://scholarworks.gsu.edu/philosophy_theses)

---

## Recommended Citation

Pecotic, Adrian, "Isolated Sacred Value Theory: An Account of Moral Conative Attitudes." Thesis, Georgia State University, 2018.  
[https://scholarworks.gsu.edu/philosophy\\_theses/241](https://scholarworks.gsu.edu/philosophy_theses/241)

This Thesis is brought to you for free and open access by the Department of Philosophy at ScholarWorks @ Georgia State University. It has been accepted for inclusion in Philosophy Theses by an authorized administrator of ScholarWorks @ Georgia State University. For more information, please contact [scholarworks@gsu.edu](mailto:scholarworks@gsu.edu).

# ISOLATED SACRED VALUE THEORY: A DECISION-PROCEDURE FOR MORAL CONATIVE ATTITUDES

by

ADRIAN PECOTIC

Under the Direction of Neil Van Leeuwen, PhD

## ABSTRACT

In this paper, I propose a novel theory of *sacred values*, which are a recently proposed type of conative attitude meant to account for religious and political actions that are incomprehensible using theories of rational choice (Atran 2016; Baron & Spranca 1997; Tetlock et. al. 2000). Sacred values are unique mental states because they encode unconditional preferences for certain privileged outcomes. I develop Isolated Sacred Value Theory by formulating two decision principles that reflect behavior in morally-relevant circumstances: the inviolability principle and the unrankability principle. Having formulated my proposal, I consider the impact of my proposal on extant theories of rational choice (e.g. von Neumann & Morgenstern 1944). Lastly, I apply ISVT to problems regarding the typical extension of sacred values and the possibility of immoral actions.

INDEX WORDS: Sacred Values, Religious and Political Action, Choice, Moral Evaluation

ISOLATED SACRED VALUE THEORY: A DECISION-PROCEDURE FOR MORAL  
CONATIVE ATTITUDES

by

Adrian Pecotic

A Thesis Submitted in Partial Fulfillment of the Requirements for the Degree of

Master of Arts

in the College of Arts and Sciences

Georgia State University

2018

Copyright by  
Adrian Ryan Pecotic  
2018

ISOLATED SACRED VALUE THEORY: A DECISION-PROCEDURE FOR MORAL  
CONATIVE ATTITUDES

by

Adrian Pecotic

Committee Chair: Neil Van Leeuwen

Committee: Daniel Weiskopf

Electronic Version Approved:

Office of Graduate Studies

College of Arts and Sciences

Georgia State University

August 2018

## **ACKNOWLEDGEMENTS**

The time in my life during which I wrote this thesis and the generosity of those surrounding me ensure that I have many people to thank for making it possible. Most of all, I must thank all those who cared for me: my family, all my friends in both Halifax and Atlanta, and especially Zach, Mara, Hyla, and Hansen. The Philosophy Department was helpful and accommodating when it most needed to be, for that I thank Eddy Nahmias. Bobby McKinley, Clare Mauney, and Daniel Weiskopf all provided welcome critiques of content and style. And, of course, I must thank my advisor, Neil Van Leeuwen, for all his guidance.

## TABLE OF CONTENTS

<b>ACKNOWLEDGEMENTS .....</b>	<b>V</b>
<b>LIST OF TABLES .....</b>	<b>VII</b>
<b>LIST OF FIGURES .....</b>	<b>VIII</b>
<b>LIST OF ABBREVIATIONS .....</b>	<b>IX</b>
<b>1 INTRODUCTION.....</b>	<b>1</b>
<b>2 ISOLATED SACRED VALUE THEORY .....</b>	<b>3</b>
<b>3 EMPIRICAL FINDINGS AND JUSTIFICATION OF ISVT.....</b>	<b>8</b>
<b>3.1 Summary of ISVT .....</b>	<b>17</b>
<b>4 CONSEQUENCES FOR RATIONAL CHOICE THEORY .....</b>	<b>20</b>
<b>5 EXTENSION OF SACRED VALUES AND THE POSSIBILITY OF IMMORAL ACTION.....</b>	<b>33</b>
<b>6 CONCLUSION .....</b>	<b>43</b>
<b>REFERENCES.....</b>	<b>45</b>
<b>APPENDIX: A FORMALIZATION OF ISVT.....</b>	<b>49</b>

**LIST OF TABLES**

Table 1 Notable Features of Sacred Values ..... 9



**LIST OF FIGURES**

Figure 1 Hyperbolic vs. Exponential Discounting.....	28
--	----

## **LIST OF ABBREVIATIONS**

ISVT: Isolated Sacred Value Theory

RCT: Rational Choice Theory

## 1 INTRODUCTION

We might imagine someone running errands one morning: carefully planning an itinerary, finding the best parking spots, and scanning each aisle for a good deal. We might imagine the same person following that unremarkable morning with actions that possess very different qualities, like going to a risky protest or engaging in civil disobedience. General theories of behavior struggle to explain the full range of human action with a few theoretical principles; one influential attempt is Rational Choice Theory (RCT), which explains action as an expression of the highest ranked of an agent's preferences (Peterson 2009; von Neumann & Morgenstern 1944). However, this analysis seems ill-fitting when applied to acts of passion, self-sacrifice, or even cooperation (Hursthouse 1991; Atran 2006; Skryms 1998). Although advancing theories with maximal scope is admirable, I contend that a more modest approach, tailoring separate theories to subsets of actions that share relevant characteristics, results in better explanations.

Religious and political actions form a cluster whose properties distinguish them from other types of action, indicating that these acts are a subset in need of a theory. Terrorists or freedom fighters taking great risks for the cause, religious congregations steadfastly defending sacred objects or beliefs, and both civil and violent disobedience of state power are three paradigmatic examples of religious and political actions. Recently, social psychologists have posited "sacred values," which are desire-like mental states with properties that preclude compromise and discourage free-riding (Atran 2016; Baron & Spranca 1997; Tetlock et. al. 2000). However, these accounts lack necessary features of a philosophical account, like integration with philosophical understandings of action and motivation.

Philosophical theories of action identify mental states as the cause of actions, casting explanations in psychological terms meant to describe actual mental processes. The standard

approach follows David Hume's notions of 'reason' and 'passion,' which correspond to beliefs and desires, respectively (Smith 1987). When an agent wants something and believes a certain action will satisfy her want, she performs the action (Davidson 1963; Smith 1987). RCT assumes a similar framework, replacing desires with preferential orderings and beliefs with subjective probability (von Neumann & Morgenstern 1944). To introduce a new type of mental state as an explanatorily-relevant factor in action, as sacred values have been, one must assess the impacts on longstanding theory. And, of course, the main task is to develop a set of principles that describe the interactions and functions of the mental states that cause action.

I accomplish both the above goals with theses concerning (1) how sacred values function as mental states and (2) the relationship between sacred values and non-sacred values. With regards to the first goal, I develop *Isolated Sacred Value Theory* (ISVT), which aims to understand the structure and function of sacred values; here, my central contention is:

- (1) Each sacred value is an isolated preference structure, which is inviolable by profane values and unrankable against other sacred values.

This proposition contains two substantive claims concerning the comparability of value-types: firstly, the inviolability principle states that agents cannot exchange a sacred outcome for non-sacred goods, like enjoyment or money; and secondly, the unrankability principle holds that the value of a sacred outcome cannot be measured on a common scale with other sacred outcomes. The appendix contains formalizations of these principles. Together, the two central postulates of ISVT suggest a moral motivational system wherein each sacred value is isolated from both profane values and other sacred values. As for the second goal, I will argue that adopting ISVT does not require abandoning RCT, only restricting the scope to exclude any choices featuring sacred values. Thus, my second thesis is:

- (2) Standard actions and sacred actions are caused by two separate competencies, with the former described by RCT and the latter by ISVT.

As will become clear, the content of a conative attitude does not determine the competency which processes it, since the same prospective state-of-affairs could be valued as sacred or profane by different agents. When an agent does not identify sacred value in any available option, one can still use RCT to understand the behavior.

In Section 1, I introduce the core postulates of Isolated Sacred Value Theory. In Section 2, I review empirical findings by social psychologists studying religious and political decision making, further defending and developing my theory (Atran 2016; Baron & Spranca 1997; Tetlock et. al. 2000). In Section 3, I consider the consequences of ISVT for our appraisal of RCT and argue that theories of high-level psychological phenomena need not correspond to actual processes in the brain nor make flawless predictions. Instead, these theories must properly demarcate the domains of choice and discover decision-making principles. Finally, in Section 4, I consider the possibility of immoral action and regularities in what is considered sacred across cultures.

## 2 ISOLATED SACRED VALUE THEORY

In this section, I will explain the basic features of ISVT, but neglect full arguments for many theoretical moves until the next section, when I review the aforementioned empirical literature on sacred values. First, I introduce the distinction between sacred and profane value. Second, I advance the two central posits of ISVT: the inviolability of sacred values by profane values and the unrankability of sacred values. Finally, I define sacred values as conative mental states with structural properties of the form  $[S > A]$  and  $[S_1 \dashv S_2]$ .

On my theory, sacred and non-sacred preferences are two different conative attitudes an agent can adopt towards a state of affairs. Emile Durkheim first introduced a psychological

understanding of the distinction between the sacred and profane in *The Elementary Forms of Religious Life* (1912). Durkheim argues that religion is a system of beliefs, rites, and practices that maintain distance between what is sacred and what is profane; he writes: “by definition, sacred beings are beings set apart. What distinguishes them is a discontinuity between them and profane beings” (1912 p. 303). Although ‘sacred’ is most often associated with religion, the term also applies to secular objects and places that inspire reverence, like the American Flag, Simon Bolivar’s sword, or Military Cemeteries. Whether religious or secular, we treat these objects and places according to rules of conduct that ensure their separation from the profane world—one is not supposed to burn a flag, nor should one litter at a Cemetery. It is the collective respect for a sacred value that affords objects, places, and ideas their aura and import, not any property of the things themselves. Profane value does not necessarily disrespect the sacred, they are simply irrelevant to sacred concerns. Non-sacred objects only become sacrilegious when illicitly elevated to the sacred plane. Throughout the essay, whenever I refer to non-sacred or profane values, I mean the sense just introduced. Behind these acts and prohibitions lie conative attitudes that afford special importance to the maintenance of sacred states of affairs.

The sorts of things people consider sacred varies considerably between times, places, and, of course, the political spectrum. Accordingly, I will argue that sacred values are a conative attitude defined not by content, but by their functional role in decision-making. So, my theory treats terrorist attacks alongside climate change denial and Colin Kaepernick’s anthem protest. Obviously, these acts are not morally equivalent, but an agent who values a sacred outcome beyond any profane encroachment is responsible for each—only the content of the mental state differs. I am concerned with describing the psychological properties of morally-relevant decisions, not with

the normative status of what is considered sacred in any context.<sup>1</sup> Since to be sacred is just to be processed in a certain way that projects sacred value into the world, anything can, in principle, become a sacred value. Therefore, I cannot give necessary and sufficient conditions for when a state of affairs is a sacred value, relying instead on functional definitions. Though I have often referred to my explanandum as ‘religious and political actions,’ this formulation merely uses content as a heuristic for identifying actions motivated by sacred values.<sup>2</sup>

I now examine the inviolability principle, which holds that a profane outcome cannot be compared with, nor chosen over, a sacred outcome. According to RCT, any choice features multiple options one must evaluate by weighing beneficial attributes against negative ones, enabling the comparison of outcomes and, eventually, an act. In contrast, then, the presence of inviolability implies that there is no preference ranking (better than, equal to, or worse than (Chang 1997)) between sacred and profane outcomes because comparing them is improper. Absent standard preferential relationships, people display unconditional commitment to objects, places, and ideas sacralized by the community. This treatment of privileged outcomes bears considerable resemblance to Joseph Raz’s “constitutive incomparability,” a relation that obtains when “the refusal to trade one option for the other is a condition of the agent's ability to successfully pursue one of his goals” (1986 p. 346). For example, only those who believe “that friendship is neither better nor worse than money, but is simply not comparable to money or other commodities, are capable of having friends” (1986 p. 352). Certain goals and roles, like prison reform or being a priest, require treating some outcomes as inviolable, blocking the sacred outcome from entering into any of the three traditional value relations with a non-sacred outcome.

---

<sup>1</sup> In fact, from a normative perspective, those committed to sacred values often commit immoral actions in their defense, as we see in ethnic conflicts with depressing regularity.

<sup>2</sup> Though, as we shall see, despite the degree of freedom one has in adopting different attitudes towards outcomes, there regularities in the sorts of things considered sacred across cultures.

The behavioral consequences of inviolability are apparent in conflicts between sacred and non-sacred values, most obviously when money is on offer for violating a sacred value. If someone has an opportunity to sell a sacred place for land development, exchanging an increase in wealth for the destruction of the site would disrespect the place itself and those to whom it is sacred. After all, if the sacred can be traded, bought, or sold, it is not appropriately distanced from everyday commerce. So, to even *consider* trading a sacred good for material benefit is to opt out of living as a proper member of the community for whom the outcome is sacred; even if this does not literally mix the sacred and profane, it is a mental act that unacceptably mingles the two types of value by treating them as comparable. This view allows for gradual change in a person's sacred commitments over the course of her life, with development based on evolving social consensus, joining or leaving a community, and personal reflection. The fact that an outcome may be considered sacred at one point, but not another, does not inhibit the detection and study of sacred values (by way of the features identified in the next section), it means that an agent's moral, religious, or political beliefs have changed between observations of behavior. In sum, inviolability holds that outcomes realizing sacred values are incomparable with profane values because respecting sacred goods is tied to a person's ability to live a certain kind of life.

Next, I explain the unrankability of outcomes realizing sacred values with different contents. People care deeply about a great many issues; one may hold sacred values about preserving natural environments, protecting LGBTQ rights, and Palestinian statehood at the same time. Forced to choose between sacred values, agents are unable to weigh different types of value and, as a result, cannot decide on a ranking of the options. Once again, none of the conventional preferential relations can obtain, but when two sacred values are at issue, we need a relation with different properties than inviolability. Unlike inviolability, there are no behavioral prescriptions



associated with unrankability, since both options are privileged relative to the non-sacred outcome. The lack of a common scale along which different values can be compared, sometimes called value incommensurability in the literature, is responsible for the unrankability of sacred values (Broome 2000; Chang 1997). One cannot form metrics of comparison because what is considered good is very different in each outcome. For example, how does one weigh the good done by protecting the Amazon rainforest against the introduction of pro-transgender legislation? The source of our inability to construct rankings is a relationship between the abstract values themselves, not the outcomes in which they are realized.

When choosing between outcomes instantiating two unrankable values, an agent will be uncertain about the choice because the values cannot be ranked. In order to rank the outcomes, the agent needs a common scale on which the concrete bearers, or units, of each sacred value can be placed. Without a third value commensurable with both sacred values, in terms of which a common scale could be constructed, there is no way to compare 1000 acres of rainforest to a particular amount of another social good. If forced to make a choice, the agent may well choose one or the other, but this behavior is not the expression of a stable preference. Rather, a whim or arbitrary tiebreaker will have made the difference instead of a proper reason for action. Later on, the same agent may well choose the other sacred outcome. In the next section, I use the small improvements argument to definitively show that no sacred option is better than the other, nor are they equal (Broome 2000; Chang 1997; Raz 1986). Unrankability isolates sacred outcomes from one another, generating uncertainty because there is no clear reason to act one way or the other.

With the inviolability and unrankability principles in view, so is the basic structure of sacred values according to ISVT, which I elucidate by introducing symbols for the new value relations (see appendix for full formalization). Inviolability requires that any profane outcome be

rejected when in competition with a sacred outcome, which we can formalize as the preference relation  $[S > A]$ . Here,  $S$  stands for any outcome that instantiates a sacred value and  $A$  stands for *any* outcome with only profane value. The relation, ' $>$ ,' establishes that sacred outcomes will be chosen over profane outcomes, no matter the amount of value in each. Importantly, though  $[S > A]$  means that the sacred value is preferred to profane values *in toto*, this is not a preference ranking in RCT's sense, for two reasons: first, rational choice theory's rankings are between concrete outcomes, not variables for any outcome instantiating only profane value (like ' $A$ '); and second, my preference structure does not include subjective probabilities. We also need a relation to symbolize unrankability, for which I use the middle term in  $[S_1 \dashv S_2]$ . This reflects the inability to measure the values of two sacred outcomes in a way that enables an appraisal in the same terms. These two postulates leave agents with, on the one hand, severe restrictions on acceptable behavior in choices involving sacred and profane outcomes, and on the other, no definitive guidance when choosing between sacred options.

### 3 EMPIRICAL FINDINGS AND JUSTIFICATION OF ISVT

In this section, I review experimental evidence from the social psychology literature on sacred values, an area of research that emerged in the late-nineties with the work of Scott Atran, Jonathon Baron, and Philip Tetlock. Several features of religious and political action, shown in the chart, describe the ways people act when the sacred is at issue, according to which we can appraise ISVT. The first two distinguishing features relate to the inviolability and unrankability principles, respectively. The third and fourth features are implied by those same principles. The final two features emerge when one examines sacred values from the level of groups, rather than individuals. I will review evidence for each, explaining how ISVT reflects these features of sacred values in a coherent manner that follows from its central postulates.

*Table 1 Notable Features of Sacred Values*

<b>Distinguishing Feature</b>	<b>Example</b>
[1] Refusal to consider taboo tradeoffs	Refusing to sell land for fracking
[2] Difficulty of tragic tradeoffs	Syria: saving human lives vs. non-interventionism
[3] Reflect on moral character of actor	Australian cricket team ball-tampering scandal
[4] Lack of responsiveness to risk	Palestinian protests at Gaza border
[5] Uniform throughout a group	Hindus and sacredness of cows
[6] Become oppositional when groups conflict	Climate Change Acceptance vs. Denial

It is appropriate to begin with the refusal of ‘taboo’ tradeoffs [1] because of its centrality to both ISVT and the research on sacred values. To accept a taboo tradeoff is just to choose a profane outcome at the expense of a sacred value (Baron & Spranca 1997). When considering a policy, like the legality of abortion, an affirmative response to “this should be [allowed or] prohibited no matter how great the benefits from allowing it [or prohibiting it]” shows that no amount of material or hedonistic value will be enough to justify choosing the profane outcome (Baron & Spranca 1997 p. 7). For example, if one believes a temple is sacred ground, no reward, no matter how absurd, will compel an agent to allow disrespect of the site. In contrast, no matter how strongly one desires a profane outcome, like eating an orange, he will pass on the opportunity if offered enough money for the orange, whether that is \$5 or \$50. If someone accepts a taboo tradeoff, it not only means the sacred has been compared to the profane, but that person actually values material outcomes more than the sacred. An agent that regularly accepts taboo tradeoffs does not sincerely possess the sacred value in question.

A very consistent finding is that agents refuse, and refuse to consider, taboo tradeoffs (Atran 2016; Baron & Spranca 1997; Tetlock et. al. 2000). For example, in one of Tetlock’s experiments, people identifying as ‘liberal’ overwhelmingly refused to trade money for “votes in

elections for political offices, the right to become a U.S. citizen, [or] the right to a jury trial” (Tetlock et. al. 2000 p. 857). Likewise, while interviewing students at Indonesian *Madrasahs*, Ginges and Atran found that “75% of participants answered ‘no’” when asked to compromise their belief in sharia law, at some schools reaching 91% of respondents (Ginges & Atran 2009 p. 17). Upon the adoption of a sacred value, taboo tradeoffs are refused in principle, suggesting that a true comparison of the options never happens (Baron & Spranca 1997 p. 7). A conventional preferential ranking is unnecessary because the unconditional endorsement of a sacred value ensures that any amount of sacred good is chosen over solely profane outcomes.

What’s more, people react with “moral outrage” and “moral cleansing” to just the offer of a taboo tradeoff, signifying an opinion that the question ought never have been asked (Tetlock et. al. 2000 p. 853). These findings are consistent with inviolability, in which any openness to trading the sacred for the profane is incompatible with the larger goals of the agent. Indeed, Tetlock adopts an explicitly Razian perspective, writing, “our commitments to other people require us to deny that we can compare certain things” (2000 p. 854). For example, suppose a company that uses embryonic stem cells in medical treatments offers a devout Catholic a job with a high salary and exceptional benefits. Of course, the Magisterium forbids any destruction of embryonic stem cells, since it is tantamount to abortion and vitiates the sanctity of life. To take the job would *certainly* forfeit his standing as a good member of the Church, but to even consider it warrants harsh censure. In these terms, the agent refuses to violate sacred goods because of restrictions that are constitutive of the pursuit of the spiritual project of living as a good Catholic.

Moving on to feature [2], we are now concerned with the unrankability of outcomes that instantiate two different sacred values. These choices, often called ‘tragic tradeoffs,’ have attracted less study than sacred-profane tradeoffs, usually as contrast conditions in studies focused

on taboo choices. However, one important finding is that tragic tradeoffs are very difficult decisions to make, relative to both standard and taboo tradeoffs. When Martin Hasselman and Carmen Tanner offered a choice between improving “flood protection” or “safety at work” in a community, subjects said the tragic tradeoff was more than twice as difficult as the taboo condition (2008 p. 54). Difficult is an expected property of tragic choices according to ISVT, since the lack of a common scale entails that no obvious answer is forthcoming. Another finding lending credence to the unrankability thesis is people’s lack of confidence in their decisions concerning sacred values, again compared to standard and taboo tradeoffs (Hasselman & Tanner 2008 p. 54). Taboo tradeoff have an obvious, morally required, choice, but tragic tradeoffs are not determined by moral rules because both outcomes realize sacred values.

However, decision difficulty does not conclusively establish that sacred values are not ranked against with one another; it is possible that the properties of tragic decisions are caused by equally desirable options. For that, one must use the small improvements argument to show that none of the three preference relations obtain between the two options, which implies there is no common scale (Broome 2000; Chang 1997; Raz 1986). For example, take an Evangelical agent confronted with two policies, one in which results in 2500 fewer abortions a year and the other in which 2500 more people become ‘born again’ in a godless country. As we have seen, the agent will not see one outcome as obviously preferable to the other, since the comparison is difficult. In fact, agents often deny the possibility or necessity of sacrificing one value for another (Baron & Spranca 1997). Thus, even if the agent does make a choice, we have reason to doubt that the agent has a stable preference for one or the other outcome. To show the agent does not see the options as equal, we must introduce a second, slightly altered choice. If one outcome is ‘sweetened,’ and the choice is now between 2500 fewer abortions and 2505 more people born again, the agent’s

indecision will persist. The extra utility from 5 new congregation members entails that, had the two options truly been equal in the first choice, the agent should choose the sweetened option without hesitation in the second. We can conclude that agents are not comparing the values instantiated in each outcome along a common scale because, if they were, sweetening's marginal increase in utility would result in a strict preference.<sup>3</sup> Therefore, we have good reason to believe that different sacred values cannot be ranked, establishing the second core property of ISVT.

Thirdly, by attending more closely to results concerning moral outrage and the judgment of others, we can improve our understanding of the ways in which decisions featuring sacred values impact one's moral standing [3]. Readers of vignettes *about* agents making taboo decisions report "moral outrage" at the fact the choice is even being considered, the degree of which increases with the length of contemplation (Tetlock et. al. 2000 p. 852). In taboo conditions, participants of the experiment judged the moral standing of the decision-maker very harshly, with some recommending punishment (Tetlock et. al. 2000). As expected, how an agent acts when in a position to choose between the sacred and profane affects their moral standing in the eyes of others. Further, people cast their moralizing eyes on themselves, with Tanner and Hasselman finding that "compared to routine trade-off scenarios, negative emotions were somewhat stronger in taboo trade-off scenarios and considerably stronger in tragic trade-off scenarios" (2008 p. 55). Aware of the moral import of decisions involving sacred values, people feel pressured and 'on the spot' because they know their moral standing is subject to the judgment of their community.

---

<sup>3</sup> Some may object by arguing that the small improvement argument is only sufficient to show that the outcomes are not *exactly* equal, leaving open the possibility they are *roughly* equal (a fourth value relation), and as such, comparable (Chang 2002). This challenge can be met by giving a 'large-improvement argument' with the same structure as its sibling (Boot 2017).

Importantly, the negative emotions that agents experience during tragic tradeoffs are tied to the difficulty of choice, with participants endorsing statements like “I was swamped with this decision” and “I was afraid to make a choice” (Tanner & Hasselman 2008, Appendix B). The lack of a traditional preferential ranking causes the behavioral instantiations of uncertainty and worry. Returning to the perspective of someone judging a decision-maker, we find that lengthy consideration of a tragic tradeoff does not merit derision, but instead, the attribution of ‘wisdom’ (Tetlock et. al. 2000 p. 858). Since these important decisions are so difficult, the absence of haste befits what is at stake. The least difficult decisions involve only profane value, since the outcomes are commensurable with one another and enter conventional preferential relations. RCT expects the ease of standard tradeoffs because the completeness axiom dictates that an agent can form a preference between any two options, but is inapplicable to taboo or tragic choices because comparison fails in each, though in different ways. Inviolability and unrankability have different behavioral consequences, but, in both cases, the agent is aware that her behavior will represent her moral character to others in the community, and, to herself.

Feature [4], the lack of responsiveness to risk, enables many of the most salient and extreme actions caused by sacred values. When agents make costly, even deadly, sacrifices for a sacred value, it seems as if the likelihood of high non-sacred costs are disregarded. This inattention is vividly apparent in suicide terrorism, where costs include hardship for the bomber’s family and, of course, his own life are nearly certain. Atran writes: “those who have become radicalized to Jihad (as measured by support for suicide actions) respond to both instrumental sticks (enemy counter-violence) and to instrumental carrots (recall the results for Palestinian refugees on taboo+ tradeoffs) with even greater support for violence” (2006 p. 4). Thus, an increase in cost, represented here as ‘counter-violence,’ increases agential willingness to commit the action; that is,

the more profane goods the agent is risking, the less the profane matters. The insensitivity to profane costs associated with sacred outcomes suggests the processes that evaluate these outcomes do not include estimations of benefits, costs, or risks. The steadfastness with which people choose sacred outcomes in the face of increasing profane costs is consistent with the presence of any amount of sacred value ‘activating’ the rules of ISVT, meaning inviolability continually precludes compromise. Since the goods available in sacred outcomes are not measured and bundled together, but simply recognized as sacrosanct, there is no ‘sum’ of value to combine with subjective probabilities in order to calculate the act’s expected utility. As such, agents cannot integrate probabilities associated with an outcome into the decision-procedure, which means agents motivated by sacred values will respond to risks differently than in profane tradeoffs.

The next feature, the uniformity of sacred values throughout a group [5], departs from the features above because it concerns the behavior of groups, rather than individuals. This requires abstracting away from individual agents and considering values held in common by groups of people. Both perspectives are necessary for understanding sacred values because individuals do not create their own sacred values; rather, we receive our ideals from our society. However, this is not a purely receptive process, but one in which each person committed to a sacred value contributes to its continued place in the community. Mechanisms for the creation and maintenance of sacred values include public rituals and demonstrations, on which Durkheim writes, “all parties—be they political, economic, or denominational—see to it that periodic conventions are held, at which their followers can renew their common faith by making a public demonstration of it together” (1912 p. 212). Holidays, political rallies, and Sunday mass are rituals with the social function of strengthening or developing shared sacred values, thereby increasing group cohesion. Testing these propositions, Hammad Sheikh and colleagues found that “the more frequently



participants took part in religious rituals, the more they considered their preferences to be sacred values” (2012 p. 114). These rituals synchronize people’s conception of the margins of the sacred world, weaving individuals into a collective with shared goals and pursuits.

Since rituals are explicitly *directed* towards an object, place, or idea, participants will adopt a sacred attitude towards the same content, committing to applying the inviolability and unrankability principles in the relevant contexts. When a group is prepared to cooperatively defend a sacred value without compromise, any challenges will be met with greater force than even the strongest lonesome defender could muster. One of Atran’s key findings is the greater the degree of “identity fusion” with a group, the more likely it is an agent engages in “extreme pro-group behavior” (2016 p. 197). Identity fusion, on Atran’s theory, occurs when “people’s collective identities become fused with their personal self-concept” as a result of emotional bonds between groups friends, family, and “kin-like” social groups (2016 p. 197). Furthermore, treating a sacred object properly “credibly signals” one’s commitment to a group, especially since the unconstrained nature of the inviolability principle ensures adherence to a sacred value is “costly to fake” (Sosis & Alcorta 2004 p. 267). Richard Sosis finds that “a higher number of costly rituals and taboos” in a commune, like a kibbutzim or collective farm, the greater the average length of survival before dissolution (Sosis & Alcorta 2004 p. 271). Thus, people devoted to the same sacred value have common ground on which to build mutual trust, solving problems of free-riding and cheating. The expectation that members of a group will act in defense of a sacred value, even when counter to their own interests, enables a level of cohesiveness not present among purely self-interested individuals; it is at this point, as Durkheim puts it, people form a “society” (1912 p. 16).

Our propensity for inter-group competition and conflict is an unfortunate feature of humanity, and when it happens, the sacred values of each group become oppositional [6], which

further drives the conflict. By ‘oppositional,’ I mean that one group’s sacred object becomes a profane object to the opposed group, and vice-versa. For example, American flags are sometimes treated as profane objects in the Middle East, to be burnt and spit upon, while more Islamophobic corners of the 50 States have had the exact same ideas about the Koran—but both sub-cultural groups perform these rituals with equal rarity. The slurs used in war to refer to the enemy ingrain the profane, disgusting, nature of the enemy into the minds of recruits (Kubrick 1987). The damage inherent in this type of violent oppositional thinking should be clear to everyone, as it has played out countless times throughout history.

The empirical evidence contains results which lend credence to the development of oppositional sacred values, though no one has studied the phenomena specifically. Atran writes, “devoted actors are most likely to commit themselves to extreme actions of parochial altruism if they perceive themselves to be under existential threat from outside groups” (2016 p. 198). Further, Sheik et. al. find the process of creating new sacred values “is accentuated by the presence of intergroup conflict,” with new values likely vilifying the other group (2012 p. 116). Sacred values are sensitive to the always changing needs of a community, especially when in competition with other groups. When open hostilities break out, the act of desecrating something considered sacred by the other group can itself assume sacred value for the actor and his neighbors. The structure of sacred values I’ve described,  $[S > A]$ , easily accommodates this feature if one considers the sacred objects of the opposed group to take the place of ‘A,’ such that its very presence is an affront to the sacred values of the group. Thus, acts against what is sacred to another group become required in taboo tradeoffs because to act otherwise is automatically an affront to one’s own values, allowing the sacred object of the other group to tarnish one’s own sacred values.

### 3.1 Summary of ISVT

My theory of sacred values can explain all the features of religious and political action just surveyed, even though they occur at different levels of analysis. The first two features, the refusal to consider taboo tradeoffs and the difficulty of tragic tradeoffs, are observations of behavioral responses to different types of choice situations. In contrast, the uniformity of sacred values amongst members of religious, class, political groups, as well as their tendency to become oppositional when these groups conflict, concern groups of actors unified by their values, not individuals. ISVT contributes to understanding both levels of analysis with the same small set of principles about the interaction of conative mental states.

ISVT introduces two new relations that obtain between outcomes in choices that implicate one's sacred values. Inviolability ( $>$ ) and unrankability ( $\nmid$ ) are rules that determine the result of comparisons based on the agent's attitude towards the outcomes. Since these relations predict an agent's choice using only attitude-type, not content, the rules generalize to any choice involving sacred values. Thus, whenever an agent is offered a taboo tradeoff involving anything taken to be sacred, she will refuse the profane reward, no matter the offer [1]. The incompatibility of social acceptance with sacrilegious action ensures the automatic rejection of taboo tradeoffs, especially since censure follows mere consideration of an offer. Similarly, without regard for content, any conflict between two different sacred values generates uncertainty [2], since the values are unrankable. Beyond this, ISVT cannot make ironclad behavioral predictions about tragic tradeoffs because their relative value is inherently undecidable, so nothing about the conative attitudes themselves provides grounds for prediction. Though this may be seen as a deficit of ISVT, I would posit that it properly reflects uncertainty present in the phenomena.

When people see or hear of how others behaved in choices involving sacred values, they judge the moral character of the actor [3], the valence of which varies depending on the type of tradeoff. As we know, taboo tradeoffs must be rejected quickly to prevent moral outrage. In contrast, agents are not negatively appraised for taking time to consider the response to a tragic tradeoff, appropriate since the unrankability gives no guidance on the right action. In both cases, the way people react is consistent with inviolability and unrankability. Likewise, ISVT expects the lack of risk responsiveness in decisions involving sacred outcomes [4] because the structure of sacred values precludes the integration of subjective probability into decision making. In taboo cases, the options are not actually compared, with the profane outcome rejected out of hand; thus, the likelihood of success in the sacred action is not relevant. And, in tragic tradeoffs, there is no common value to use in the calculation of expected utilities for the outcomes, so probabilities do nothing to resolve unrankability.

Finally, ISVT connects individual mental states to social conditions using the link between public rituals and the sacrilization of new values (Durkheim 1912; Sheik et. al. 2012). The collective behavior of a community unifies the contents of individual's conative attitudes [5] by directing everyone's attention towards a common object, place, or even concept, creating a shared sense of the significance of outcomes instantiating that value. A group of individuals needs a shared set of sacred principles to coalesce into a group capable of large scale cooperation; they need a society, otherwise they remain autonomous utility maximizers. The structure of sacred values aids our understanding of conflict between groups, which each possess their own sacred values. We can symbolize oppositional sacred values as  $[S^i > S^{ii}]$ , with the primes distinguishing values that belong to different groups. This structure implies that the presence of the offending

group's value,  $SV^{ii}$ , is an affront to the other groups sacred value,  $SV^i$ ; this, of course, motivates single-prime people to remove any offending sacred objects belonging to the double-primes.

Before ending this section, I wish to raise and respond to a counter-proposal. The central importance of group membership in my account of sacred values suggests that we may be able to explain sacred actions in terms of desires to belong to a group, without mentioning sacred values at all. On this proposal, a sacred acts are caused by higher-level preference for remaining in the group without any attachment to the 'sacred value' purportedly instantiated in the outcome. If, as I have claimed, a precondition of belonging to a group is respecting their sacred objects, it seems possible to interpret any action in defense of those objects in a way that affords great value to remaining in the group and no value at all to the sacred objects themselves. Although initially credible, this proposal does not capture (1) private sacred actions or (2) the externalization of sacred value.

First of all, if agents only value sacred objects to the extent they contribute to one's social status, it is difficult to see why an agent would ever privately perform sacred acts. Once one has adopted the attitude that an object is sacred and must be treated according to certain rules, the agent will continue to apply those rules even when no one is looking, as we do when persisting with effortful sacred practices, like the proper method of prayer or folding a flag, even when no one is looking. The independence with which agent's engage with their sacred duties shows that they value the outcomes in themselves, not as a means to another end. Relatedly, the present proposal does not explain the way in which we externalize sacred value, projecting it onto states of affairs but not others. Many important sacred objects have an aura, the source of which is "collective effervescence" that affords the "sacred forces... special qualities that the profane have not," like an individual's "idea.. that there is an infinite power outside him to which he is subject" (Durkheim

p. 82). Kyle Stanford has recently taken up these themes, arguing that externalization establishes a “crucial connection between our own motivation to conform to any given distinctively moral norm of behavior and the extent to which we demand that others conform to the same norm” (2018 p. 2). The fact that we externalize sacred values raises our commitments from standard preferences to “distinctively moral obligations,” which augments group cohesion to a greater extent than shared commitment to merely instrumental values (Stanford 2018 p. 8). The proposal that we value sacred objects only as a means to the end of group membership is incompatible with our independent respect for the sacred and our externalization of moral obligations.

#### **4 CONSEQUENCES FOR RATIONAL CHOICE THEORY**

The existence of a competency solely responsible for processing decisions involving sacred values means that what was thought to be a unitary target of inquiry, decision-making, actually contains two targets, sacred decision-making and non-sacred decision-making. Discoveries that a natural phenomenon possesses only illusory underlying unity are not uncommon in science, occurring when we find out there are four species of giraffe or identify etiologically-distinct illnesses with identical symptomology. A successful theory has a scope that is sensitive to differences between distinct types of phenomena that merely appear similar, ensuring that theoretical entities covering heterogeneous objects do not confuse the interpretation of experimental results. Importantly, the cluster of human behavior within the scope of ISVT more or less corresponds to a pre-theatrically identifiable class of actions centered around religion and politics. These considerations raise the question of how to theorize about behavior outside the scope of ISVT; namely, actions emerging from choices between two profane outcomes.

While the contents of sacred values usually concern exceptional religious and political states of affairs, non-sacred values generally apply to the mundane outcomes that occupy most of

our attention. Countless inconsequential decisions must be made on a daily basis; whether to drive to work or take the train, whether to employ one investment strategy rather than the other, or what to cook for dinner. When people are making these decisions they are often, though not always, attempting to maximize or economize a finite resource such as time, money, or effort. One considers the choice of automobile instead of light rail ‘correct’ if the time spent commuting via automobile ends up being less than the light rail—assuming the agent is not an environmentalist, whose decision will involve ISVT. Initially, it seems reasonable to preserve RCT for explaining this class of actions because agential goals align with the decision procedure’s guarantee that a rational agent will experience the greatest benefit in the long run. In fact, restricting RCT’s scope to only profane matters improves its descriptive adequacy, given that moral choices have always resisted purely rational terms. This move rules out the most ardent interpretations of RCT (which claim its decision procedure to be operative across *all* choices), but enables a theorist to narrow the explanatory scope to a specific cluster of actions, as was done in ISVT.

However, many scholars would justifiably question the use of RCT in light of recent interest in non-rational understandings of decision-making like prospect theory and behavioral economics more generally (Kahneman & Tversky 1979; Ainsley 2002; Gigerenzer & Selten 2002). Although I will draw the reader’s attention to both empirical and conceptual reasons to be skeptical of RCT’s demise in this section, I first wish to make clear the reasons I take on the defense of such an unpopular theory in the present context. Further, I must note that because I consider ISVT to be an isolated competency, its truth or falsehood does not depend on the status of RCT, meaning the risks to my most substantive proposals are minimal (as are the benefits, one may think). In fact, this discussion will improve our understanding of ISVT for two reasons. Firstly, RCT’s decision procedure uses formalized and idealized principles to model decision-making while

remaining agnostic towards content, which represents a similar approach to ISVT, albeit with different rules. The similarity in theoretical features enables a useful contrast between ISVT and a decision procedure with different rules and functions that processes everyday choices. Secondly, accounts of multiple types of decision-making afford insight into the shifts in behavior and demeanor across moments in which one or the other competency is operative. A person must split his attention between the social concerns of sacred values and the more individualistic concerns of material preferences depending on the context and outcomes involved.

To make good on these promises, I first will briefly present the important concepts necessary for understanding RCT, as well as some common criticisms. I then introduce Chomsky's performance/competence distinction to characterize seemingly adverse experimental results as performance errors. Following this, I provide two positive arguments for RCT: the first focuses on the way in which agents correct performance errors upon further consideration, the second on the conceptual necessity of assuming rationality. Finally, I discuss the implications of a motivational system split between two separate competencies that produce behavior using different rules of practical reasoning.

The central contention of RCT is that, if an agent's preferences accord with a few simple axioms, they can be described by a 'utility function' that ranks all possible outcomes available to the agent in terms of the amount of utility, which can be fleshed out in terms of pleasure, money, or the good. Defined in this way, preferences integrate judgments of value with beliefs about the likelihood of events, resulting in a consistent ranking of outcomes that represents the expected utility of each possible action. An agent who always performs the action with the highest expected value is 'rational.' A rational agent will experience more utility, in the long run, than someone who does not maximize expected utility (see von Neumann and Morgenstein (1944) for a



mathematical proof). Thus, the image of human nature motivating RCT is of people attempting to do the best they can, whatever their circumstances.

In order to guarantee the greatest reward possible, RCT imposes axiomatic restrictions on the preference set. While there are a few variants of RCT, each with slightly different axioms, none of these divergences are relevant (Jeffery 1965; Savage 1954; von Neumann & Morgenstern 1944). No matter the system, the two most important axioms are transitivity and completeness. The transitivity axiom limits an agent's ranking of more than two options, stating: 'if  $A > B$  and  $B > C$ , then  $A > C$ .' This rule ensures preferences are consistent with one another, eliminating preferential cycles that make the agent liable to be money-pumped. Completeness requires that a rational agent can form a preference between any two options or is indifferent between them. More formally, the completeness axiom states: 'given options A and B, either  $A > B$ ,  $B > A$ , or  $A = B$ .' This assumption is required to ensure that the agent experiences the greatest expected utility possible because it guarantees that no option outside the calculus has a greater utility than those included. It further "entails that any two options, no matter how disparate, can be compared," which means an agent is able to rank situations with very different *types* of good (Peterson 2009). This axiom conflicts with inviolability and unrankability, since both preclude comparisons between the types of value represented in their outcomes

Despite RCT's staying power, the approach has faced a number of challenges, most of which concern behaviors that contradict the axioms. Behavioral economics emerged from the influential work of Amos Tversky and Daniel Kahneman, who identified 'heuristics and biases' that cause violations of basic rules of logic, probability, and decision-making. For some examples, Tversky challenges the transitivity axiom by finding that some experimental subjects exhibited intransitivity over series of choices lasting five weeks when the options involved multiple

attributes, like selecting college applicants for admission (1969). Similarly, both Tversky and Kahneman found that differential framing of identical options will affect an experimental subject's rankings. The famous example is the 'Asian Disease Problem,' which presents a choice between healthcare policies that are framed in terms of lives-saved or number of deaths. Presenting the policy in terms of deaths moved people to choose more risky policies, presumably to avoid the salient costs of death (Tversky & Kahneman 1981). Recently, George Ainsle showed we discount future rewards in a hyperbolic curve that diverges from the exponential discounting curve implied by RCT. He argues that people discount future rewards too greatly (even when the reward is quite soon), generating irrational patterns of choice like addiction and procrastination (Ainsle 2002). These three findings are representative, but certainly not exhaustive, of experiments indicating that actual behavior is irrational according to RCT's axioms, seeming to render the theory untenable because of anomalous results.

However, there is an alternative reading of the non-rational behavior which distinguishes between competence and performance, first introduced by Noam Chomsky. A competency is a cognitive system able to perform some useful function, governed by rules that determine the output of the system (1965). A competency enables some advantageous behavior or mental processes, such as Chomsky's universal grammar that enables the existence of language. Chomsky's theory of grammar is an idealized system of recursive rules that structure language, capable of generating an infinite number of sentences. Likewise, RCT's axioms describe a competence by which agents evaluate possible options and make choices in a way that ensures the most benefit. The theory presents a logic of decision-making that necessarily ignores psychological limitations for the sake of building a description that mathematically ensures the highest possible reward.

Chomsky's grammar and RCT are both idealized theories in the sense that their full expression would assume cognitive capacities beyond ours. RCT does not introduce restrictions on the number of options available to an agent at any one time, which, for anyone, includes a tremendous number of actions. Recall, the completeness axiom dictates that *every* option be included in the utility function, requiring the ability to process a nearly infinite amount of information. In actuality, people could not possibly consider *all* available options, limiting consideration to relatively few outcomes. Chomsky's grammar is also capable of generating more information than a person can handle, which he illustrates with unintelligible, yet grammatically correct, sentences with many nested clauses, such as, "the man who the boy who students recognized pointed out is a friend of mine" (1965 p. 11). When rules are open ended or recursive, it is impossible for limited cognitive agents to handle very long chains of options or sentences. However, despite its failure to achieve literal truth, RCT may still improve our understanding of choice by making accurate predictions and demarcating a particular class of actions with the function of ensuring an agent derives the most benefit from her circumstances.

Since Chomsky's universal grammar and RCT are both idealized descriptions of a competency's structure, some behavior will deviate from theoretical expectations—failures called 'performance errors.' Chomsky argues that there is a "fundamental distinction between competence (the speaker-hearer's knowledge of his language) and performance (the actual use of language in concrete situations)" (1965 p. 4). The theory of a competency must bracket some of actual behavior when it obscures the patterns visible without noise caused by theoretically "irrelevant conditions [such] as memory limitations, distractions, shifts of attention and interest, and errors" (Chomsky 1965 p. 3). When someone stops and starts in the middle of a sentence, incites disagreement between a subject and a verb, or leaves a fragment unfinished, we do not

think them completely ignorant of grammatical rules, but assume the particular performance was in error. Similarly, an agent who uses the representativeness heuristic is not irredeemably irrational, but commits a performance error relative to ideal expectations.

However, Chomsky mentions some causes of performance errors that do not seem to be connected to idealization, such as “shifts in attention and interest,” but to the interaction of the competency in question with other systems in the brain (Chomsky 1965 p. 3). Competencies are enmeshed in complex networks of systems with their own functions, enabling other systems to causally influence behavior. If an angry person acts contrary to rational expectations (Hursthouse 1991), it is not immediately obvious that this counts as evidence against a theory of, specifically and only, *choice*; had the emotional system not interfered, the agent would have acted differently. The complexity and multiplicity of connections between different cognitive systems leave cognitive scientists with no choice but to abstract away from much of it. One introduces abstraction into the theory, when necessary, by simplifying the effects of external systems or ignoring them altogether. For example, ISVT’s treatment of the emotional resonance of sacred values is highly abstract, discussed only in terms like ‘moral outrage’ that do not differentiate between specific emotions, each of which may influence decision-making uniquely.

These theories of choice, RCT and ISVT, are not meant to be models that perfectly map onto actual structures and processes in the brain that produce behaviors. Rather, they are highly idealized and abstract theories constructed for the purpose of producing accurate predictions, demarcating domains of study, and establishing coherent functional descriptions of systems that are meant to aid human conduct in some way. Although we do not live in a world, nor are we the sort of beings, that allows for perfect agreement with the stringent standards set by RCT, we are nonetheless subject to the rational norms that generally structure our behavior. Before moving on,

I wish to introduce a further source of what appear to be performance errors, which is implied by the above meta-theoretical commitments. It is a live possibility that other domains of human action exist that neither ISVT nor RCT have the conceptual resources to explain, requiring a dedicated theory. I do not think, for instance, that decisions about romantic matters are made using solely rational faculties. If distinct clusters of choice-phenomena require splitting for descriptive adequacy, theories with improper scope can expect confused experimental results.

I now consider whether or not my construal of performance errors can aid RCT in responding to criticisms grounded in experimental results that show actual behavior to be incompatible with RCT. I begin with the observation that people accept their mistakes and adjust accordingly when realizing an error in practical or statistical reasoning. The tendency to self-correct conclusively settles the dilemma between characterizing an anomalous result as a performance error or evidence of the falsehood of the theory, in favor of the former. If people are able to correct their errors without assistance, there must be a capacity that facilitates the correction in the direction that accords with rational expectations. Without a system able to perform the necessary inferences and calculations, one would expect corrective adjustments to vary randomly rather than towards rational expectations. As my two examples will show, people, without direction, modify their choices and estimations of probability to accord with rational expectations in a range of conditions. Therefore, we must posit a competency able to perform the decision procedure described by RCT to account for our corrections of performance errors. Interestingly, there is a marked contrast between the ease of corrections in RCT's domain and the intransigence of devoted actors confronted with challenges to the propriety of their sacred actions; in the latter case, rational suggestions would be rejected out of hand because rational standards do not apply to sacred values. The acceptance of mistakes and subsequent

revisions mean that we take ourselves to be acting in accordance with RCT, even if mistakes sometimes cause deviations.



*Figure 1 Hyperbolic vs. Exponential Discounting*

An important application of RCT is in choices about the future, with the theory mandating a gradual reduction in value to accompany increasingly delayed gratification. If one must wait to acquire something, it is reasonable to discount its value in proportion to the amount of time one must wait. Exponential discounting curves represent the rational expectation of our evaluations of future rewards, with shorter delays reflected in the valuation of the outcome (see #1 in the graph). Recall, however, that Ainslie (2002) argues that people discount future rewards far too much until just before the good becomes available, when the values rises very steeply (#2 in the graph). Andre Hofmeyr, Don Ross, and colleagues, provide evidence that cigarette smokers are more likely to make choices in accordance with an exponential discount curve in two conditions: first, when the salience of future rewards are heightened; and second, when subjects think about their decision as part of a ‘sequence’ (2010). These revisions toward rational expectations show that hyperbolic temporal discounting is not ubiquitous and that people possess the capacity for discounting and in a rational manner. Importantly, the corrections are not deferrals to experts pointing out mistakes,

but the result of increased attention to features of the circumstances relevant to RCT's decision procedure.

In the statistical domain, Richard Nisbett and colleagues performed experiments on gambling behavior, with manipulations that made chance more salient and induced awareness of the range of possible outcomes. In both conditions, subjects were more responsive to "statistical considerations having to do with the adequacy of the sample," meaning they did not overgeneralize (i.e. use the representativeness heuristic (Kahneman & Tversky 1974)) to the same degree as control groups (Nisbett et. al. 1983 p.353). Once again, that one does not need explicit instruction to self-correct implies that people can act in accordance with rational prescriptions. The *willingness* to self-correct, in distinction from the direction of correction, implies that people believe that they should be following rational expectations when acting.

Of course, this treatment of two studies showing that people self-correct for violations of RCT is not sufficient to invalidate many years of research on non-rational behavior. They do, however, establish two crucial propositions that must be reconciled with the results of behavioral economics: we have the capacity to act according to rational principles and we believe that we should be acting according to rational principles. In addition, despite the volume of studies that show irrationality, articles that lend credence to RCT continue to be published; for example, in the contexts of transitivity and framing effects (De Martino et. al. 2006; Regenwetter & Dana 2011). So, we are faced with a confusing set of experimental results that may include the operation of multiple competencies, performance errors that only appear to be anomalous results, and contraindications of rationality and irrationality seem to shift with context. Responding to this situation requires theories with proper scope that can represent, in an abstract and idealized manner, the functional properties of choices within the domain of the theory.

In my second argument, I will claim that the assumption of general, albeit imperfect, rationality furnishes theorists with a purposive explanation of many choices between two profane values. Intuitively, RCT's decision procedure seems to be appropriate explanatory paradigm for a great many cases, from choices about the best route through traffic to searching for deals at the grocery. We make many purely instrumental decisions, most often without even noticing, that are accurately described as a process of weighing each outcome's utility, considering the subjective probability of each, and choosing the highest ranked. In non-moral circumstances, people do not often pass up the opportunity to save a little time, money, or effort at no cost to themselves, conforming to the basic picture of human motivation RCT proposes. The idea that agents are generally working to maximize the benefits available in a situation immediately suggests ways of moving forward in an action-explanation, allowing inferences from behavior to underlying preferences or from gambling to attitudes about risk.

My perspective shares much with Jonathan Cohen's treatment of psychological studies that purport to show widespread human irrationality, wherein he writes: "our fellow humans have to be attributed a competence for reasoning validly, and this provides the backcloth against which we can study defects in their actual performance" (1981 p. 318). The constitutive rationality thesis holds that, without the assumption of rationality, our choices lack a *telos* to give the interpretation a starting point (Davidson 1967; Cohen 1981). Of course, however, neither Cohen nor I believe that performance errors never occur (or, in my case, according to rules associated with different competencies as well). Indeed, Cohen argues, correctly, that "allegations of defects in performance need to be carefully scrutinized," especially when systemic and pervasive, because such errors may reveal important features of human psychology in the relevant domain (1981 p. 318). Using the conceptual resources available from a priori rational principles, a theorist can,



with enough data, distinguish between a performance error and operation of a separate competency. If a cluster of anomalous choices seem to possess a common purpose distinct from the utility maximization of RCT, they belong to a hitherto unknown faculty with the function of realizing that purpose. Without adopting such a strategy, it would have been impossible to establish the existence of ISVT. However, if the choices lack a pattern that corresponds to some useful faculty, they are mere performances errors.

Just as the assumption of basic rationality enables common sense interpretation of intentional actions, it also underlies psychological research into the phenomenon—even some of those experiments critical of RCT. For example, when Tversky and Kahneman investigate ways in which our cognitive machinery uses heuristics to approximate, rather than flawlessly compute, the subjective probabilities RCT requires, they do not reject the premise that subjective probabilities are involved in choice, instead introducing piecemeal modifications to reflect observations of behavior (1983). Granted, behavioral economics performs better than RCT in some experimental contexts, like predicting when framing effects will generate intransitive choices, but the theory does not explain the overall pattern of a choice domain (Tversky & Kahneman 1981). However, without a background understanding of the phenomena akin to the constitutive rationality associated with RCT, behavioral economists cannot reliably draw distinctions between normal functioning, performance errors, and the operation of different competencies because, absent a priori conceptualizations, all observations are of performances equally representative of the cognitive system. Psychological theories, especially of high-level capacities like choice, must be guided by a pre-theoretical conceptual understanding of the phenomena, as when conceptually-based features of religious and political action enabled the identification of a new domain of choice.

In closing, the two competencies I have described using RCT and ISVT are mental processes that instill dispositions to behave in a manner prescribed by the rules governing preferential relations between outcomes instantiating different types of value. The presence of rules governing the choice of actions enables each of the competencies to cause an agent to act towards general goals implicit in the rules. ISVT promotes pro-social behavior in the service of shared sacred values, thereby binding a group of actors into a collective, more able to unite without free-riders or cowards. RCT, on the other hand, pushes an agent to maximize material goods, whether the particular good is time, money, or pleasure. Both ways of behaving are necessary in different contexts of one's life. Sometimes, one furiously checks boxes off a to-do list, using the sound instrumental reasoning implicit in RCT's decision procedure. At other times, however, one must work with others to defend something sacred, so the disposition to refuse taboo tradeoff enhances everyone's commitment. Deployed in their proper contexts, these two modes of action provide a coherence to our pursuits that would be impossible without the psychological systems underlying choice.

The differential features of the patterns of action caused by the two competencies are reflected in of the properties of the preferential relations available in each. If two sacred options are involved in a choice they are unrankable, while if only one outcome is sacred, the profane cannot violate it and must be rejected. Neither inviolability nor unrankability can obtain between two profane outcomes—our most common choices by far; instead, the three conventional preferential relation are available. Our decision procedures push us towards the ends implicitly in these relations, as when continued refusals of taboo tradeoffs create a better integrated community, or when the calculation of expected utility nets the greatest possible return on an investment. When

people constantly make the same sorts of decisions in the same sorts of circumstances, functions emerge out of the patterns.

## **5 Extension of Sacred Values and The Possibility of Immoral Action**

In this final section, I wish to raise two outstanding questions that an theory of sacred values should be able to answer. First, I wish to examine the reasons that sacred values seem to cluster into familiar categories across cultures, like the presence of symbolic objects and objects tied to the interests of the group. The presence of coherent themes is puzzling because members of a community can, in principle, sacralize absolutely anything. Second, although RCT and ISVT are idealized and abstract theories, we have only examined performance errors in everyday choice situations. This absence seems pressing given the prevalence of immorality, society's label for moral performance errors.

The first issue I address are the regularities in the extension of what is considered sacred across different communities, requiring a position on whether basic moral principles are sacred values. Is a disposition to help one's mother or husband, over some stranger, a behavioral instantiation of a sacred value encoding the imperative to value the well-being of loved ones over strangers? Again, the immediate and surface-level reply is that anything, in principle, can become a sacred value, a consequence of my position functional roles, not the contents, of mental states determine whether they project sacred or profane value. If communities did not sacralize the same sorts of things, my shorthand of 'religious and political actions' would not have been at all informative. Sacred values form three clusters within each community: (1) the same basic rules prohibiting violence and cheating, (2) an assortment of religious or secular symbolic objects or ideas, and (3) the vulnerable interests of the group.

When evaluating the place of a basic moral principle, say, against murder or cheating, in a mental architecture that includes sacred values, one notices that aberrant behavior, like murder or theft, is essentially socially coercive. As such, these simple, universal, moral principles seem like they must be required to even start the process of forming a *society* united by more specific sacred values. Whenever one starts hanging around with a new group of people, it is never mentioned that the group has a sacred value prohibiting stealing from one another, but if you are caught stealing from friends, you won't have those friends for very much longer (and even thieves have people they trust). We should consider 'not stealing from the group' to be a sacred value because it is a constitutive part of any stable group, which is also why we could consider it to be one of the few truly universal sacred values.

In fact, as the farmer's and prisoner's dilemmas show, acting with only one's own interests in mind can, in certain circumstances, lead an agent to miss out on collaborative projects, doing worse than they could have by cooperating (Skyrms 1998). Temporarily suspending one's own interests allows greater collective benefits, but suspicion of others and one's own temptations to cheat makes cooperation difficult. From a purely individualistic perspective, one chooses an option with less subjective utility every time one decides not to steal something when detection is very unlikely. Cast in this light, these are small, habitual, refusals to tradeoff sacred goods for material ones. The existence of a competency for processing sacred values inhibits the over-application of rational principles that would harm our cooperative abilities. It usually doesn't occur to us that we have done anything morally laudatory when resisting a momentary urge to steal something cool (but someone who does looks like a jerk), but this is because these unspoken sacred values are so unexceptional they regulate our behavior below awareness.

Secondly, groups require more parochial sacred values in order to differentiate themselves from other groups and provide an identity for their members. The proper treatment a symbolic object, whether a flag, Roman Eagle Standard, or Babylonian Idol, is important for group dynamics because the “sacredness [of the symbolic object] stems from one cause: it is a material representation of the clan” (Durkheim 1912 p. 124). The arbitrary names and symbols provide insignia around which individuals conceptualize their social identities, in other cases, the identities of a group organize around ideas, like Marxism or Wahhabism. Whatever the content of a group’s sacred commitments, whatever their symbols, the main social function of sacred values is accomplished when the collective *itself* sacralized. We can now see, more explicitly, the way in which sacred and profane outcomes are *constitutively* incomparable with each other. Recall, Raz argues that refusing to even compare, let alone trade, sacred and profane outcomes is a necessary condition for genuinely engaging in some pursuits; his example runs, “only those who hold the view that friendship... is simply not comparable to money or other commodities *are capable of having friends*” (1988 p. 352). Similarly, if a sacred value is disrespected, the offending agent has shown herself to be incapable of belonging to the group which is constituted by that sacred value.

In contrast with basic moral principles, these sacred values promote agential self-identification with the core beliefs and values of a group. ‘*Everyone against stealing!*’ is a poor rallying cry, but controversial values draw new members, whether the cause is LGBTQ rights or abortion rights. Since these values endorse more specific objects than universal moral principles, they are more contentious in intra-group discourse because members of opposing groups will not agree on them, as they do about the wrongness of killing. In India, for example, differences in sacred dietary prohibitions are exceptionally contentious, with riots and violence erupting over improper treatment of cows or pigs. This case shows that the development of oppositional values

(features [6]) usually operates on sacred objects and ideas in this class. The very same unique qualities of a sacred object that motivate people to identify with a group are useful for inciting hatred of it among others.

Thirdly, even more changeable and specific sacred values arise when circumstances call for co-operation on a specific issue. Sacred values certainly respond to historical conditions over long periods of time, driven by factors like the changes to a religion's status (an Emperor converts), economic conditions (the slow decline of West Virginia coal country), and political change (revolutions). Although the long-term evolution of value systems is a fascinating topic, the changes I am concerned with are more short-term. For example, Mostafa Dehghani (2010) finds that considerable numbers of Iranians have sacralized the development of nuclear weapons, a movement only possible since the program began to generate controversy in the late 1990s. The presence of a group of people holding the attainment of nuclear weapons greatly facilitates the defense of that project, since each person with that value must act to the extent of their capacities. Moreover, the sacred attachment to nuclear weapons was strengthened by perceived threats towards Iran from the Western world, which resonates with findings that "intergroup conflicts" stimulate the creation of new sacred values (Sheik et. al 2012).

Overall, the rough taxonomy I have composed uses an organizing principle akin to Ludwig Wittgenstein's analogy between riverbeds and credential attitudes, where "the bank of that river consists partly of hard rock, subject to no alteration or only to an imperceptible one, partly of sand, which now in one place now in another gets washed away" (1969 §99). Important, difficult to doubt, beliefs—like the existence of one's own hands—are the most permeant and basic, forming the bedrock of the river that determines the possibilities of less certain, more changeable beliefs. The universal sacred values against violence and cheating are constitutive of cooperation *in*

*general*, and as such, they are necessary for the development of groups in the first place, forming the bedrock of our moral cognition. Next, we have the more salient sacred values we've focused on play a crucial roles in constituting groups and sub-groups. These sandier values are more nimble and parochial, better suited for defining the identity of the group because of, rather than despite, their idiosyncrasies. These values must be somewhat stable so that the group can continue to rally around them. Finally, sacred values that arise in response to current conditions allow groups to coordinate behavior to defend their interests. The import of these values rises and falls with circumstance, so they represent the least permeant and least important values to the long-term cohesion of a group (though they can be crucial for survival in the short term).

Moving on to my second topic, it may be difficult to see how immorality is even possible for agents following ISVT's decision procedure, since the absolute character of inviolability never allows choice of the profane. Strictly speaking, the agent should never even consider the taboo option, leaving no room for sacrilegious action. Of course, one could label these actions as performance errors and gesture at many of the same factors that affect the efficacy of RCT; however, without further illumination of the conditions that moderate performance, the move is an ad-hoc elimination of adverse data. To avoid this, I provide three explanations of different types of performance error that cause immoral behavior: (1) framing a sacred outcome as solely profane, (2) framing a profane outcome as containing sacred value, and (3) knowingly immoral behavior when personal costs are too high.

First, however, I introduce a distinction between two phenomena that both appear relevant to the discussion of immorality, but one only deceptively so. On ISVT, there are two possible appraisals of an act that violates a sacred value: first, the actor may have violated a sacred value held by an observers, but lack that particular value herself; and second, the actor may have violated

one of her own values. Only the second set of cases are instances of the problematic immoral actions we are considering here, since the absolute structure of the sacred value only affects those who hold it. Someone who does not consider a flag to be a sacred object is not subject to prohibitions regarding the proper treatment of the patterned swathe—even if an observer may find the repurposing as a cleaning rag to be distasteful. In contrast, if the agent herself attaches sacred value to the flag's dignity, she commits an immoral action by her own standards.

The second type of cases, the truly sacrilegious actions relevant here, raise a dilemma in our interpretation of the behavior: either the agent genuinely holds  $[S > A]$  but somehow acts otherwise or the person is engaged in some form of moral posturing. The behavior may be intended to express the 'normatively correct' response to the moral problem, avoiding social sanction and negative self-judgment. In the real world, money on the table, agents would compromise their moral values for a great enough payoff, meaning that our moral convictions are better described as "pseudo-sacred values" (Tetlock et. al. 2017 p. 1).

I contend that, even as sinners, people are genuinely committed to their sacred objects, places, and ideas; of course, some take sacred obligations more seriously than others. Though a greedy undergraduate who reflexively endorsed the sacredness of natural environments may, in reality, sell the land for cashable check, a committed environmentalist would not. Likewise, religious congregations adopt attitudes of varying severity towards moral principles like birth control, gay marriage, or women's rights; and, within each congregation, some members take their religious values more seriously than others. Atran's work, which features data from interviews and surveys of both captured ISIS fighters and members of the Peshmerga (a Kurdish defense force), demonstrates that some agents pursue sacred value at any cost (2016). We cannot consider sacred values to be merely 'pseudo-sacred,' nor can we deny it appears the inviolability principle



is less *absolute* than it seems. Inviolability as presented in ISVT, can never be violated, but this is an idealization of our actual decision principles that reflects the conceptual underpinnings of the moral domain. Thus, meeting the challenge of immoral action requires a positive account of the factors that cause humans to commit immoral performance errors, of which I think there are three types.

The first type of immoral action occurs when an agent frames a sacred outcome as a profane outcome in order to use the RCT competency, rather than ISVT, to make the decision. In effect, this is a denial that the outcome possesses any sacred value, even though one has, in the past, explicitly endorsed similar outcomes as sacred. A change in attitude-type means that what was a sacred outcome can now be treated as a profane outcome, able to enter into standard tradeoffs where either option is available for choice, unlike the prescribed results of taboo tradeoffs. What was the obligatory choice becomes just one option, ranked according its subjective utility. For example, a person might believe that trash ought be aggressively minimized and know that takeaway food produces an egregious amount of unrecyclable waste, but neglect to consider the environmental impacts when in need of a quick meal on a busy night. Usually, when an immoral outcome is enticing enough to motivate obscuring of one's sacred duties, it contains enough profane value to overwhelm the formerly-sacred option when it is stripped of the automatic privilege sacred value endows, which happens when our environmentalist focuses on his hunger and ignores environmental impacts. Though these actions are incongruent with stated moral positions, the agent does not completely abandon a sacred values, but merely neglects to recognize their relevance in a particular situation. One does not need to ignore the presence of a sacred value consciously, these processes likely rely on sub-personal capacities like attentional control or selective memory. Likely, the semi-intentional forgetting of sacred value introduces some

cognitive dissonance in the agent, which suggests that, in these moral performance errors, agents will correct in the direction mandated by ISVT—just as they corrected towards RCT in errors of reasoning and statistics. Embarrassment and contrition will be more prominent in these corrections, I believe, because the failures implicate the moral integrity of the inattentive sinner.

The second type of immoral action also involves a misapplication of attitude-type to outcome; this time, however, it is the profane outcome that is sacralized. Paired with an outcome realizing a legitimate sacred value, the newly sacralized outcome is unrankable, as expected in tragic tradeoffs. Since neither option is stably preferred to the other, the agent can choose the immoral action with some arbitrary justification for the necessity of making a choice. Tetlock et. al. identify this mental maneuver, using cases about the costs of healthcare decisions to demonstrate that “the failure to save a child’s life ceases to be a taboo money-vs.-life trade-off and becomes a tragic life-vs.-life trade-off when we can redirect the resources to save more lives elsewhere” (2017 p. 7). If a hospital administrator concerned with profits is able to frame the choice as a tragic tradeoff, she is not bound by the demands of inviolability, which requires immediate rejection of the monies. Hence, the agent can make an immoral choice because unrankability allows for the choice of either outcome, albeit without a preferential ranking that justifies the choice. The presentation of a profane outcome as a sacred one is sometimes less arresting than passing off the sacred as the profane, but the particulars of outcomes under consideration will dictate which strategy is least cognitively dissonant.

Thirdly, I argue that ISVT is compatible with agents making immoral choices in full knowledge of the types of value instantiated in the decision, when sacred value is too apparent to ignore. These are the types of actions that motivated the analysis of immorality, since the choice of a profane outcome seems to contradict  $[S > A]$ . For example, agents sometimes make public

spectacles of their respect for a sacred value, but privately violate them in obvious ways. In these cases, an agent uses RCT in a decision that includes a sacred outcome, thereby initiating a taboo comparison and paying the social price. Tim Murphy, a disgraced former politician, is a guiding example, having recently featured in articles with headlines like, “Pro-Life Congressman Adopts Sensible ‘Abortion for Mistresses Only’ Stance” (Ryan 2017). The clarity of Murphy’s decision—he was asking someone to get an abortion—means that he cannot have used either mode of immorality explained above; he must have consciously chosen to act against his sacred beliefs.

In such cases, the agent decides the personal costs of sacred obligations are so extreme that damaging one’s moral standing in the community is worth it for avoiding catastrophic personal costs. Tim Murphy never *forgot* the content of his votes and speeches, nor his fervent public action in support of Christian values, but when urging his mistress to get an abortion, he was considering the impact of a child outside of wedlock for his life as an individual. The suddenness, magnitude, and importance of the personal and sacred values at play in these choices may impede the proper functioning of the sacred value system, overwhelming one’s ability to be steadfast in refusing to violate the sacred. When large amounts of money are involved in a choice, selfishness can motivate a taboo acts. Likewise, a sacred act that requires daring or courage may be commissioned by those willing to accept great personal cost out of loyalty to a group, but not everyone possesses the constitution of a martyr. Many of our terms of censure—like greed, cowardice, or moral blindness—seem to might pick out different types of moral performance errors.

Recall, one ceases to be an accepted member of the community the moment one compares a sacred outcome with a profane outcome instead of rejecting the tradeoff out of hand.<sup>4</sup> When an

---

<sup>4</sup> An interesting related phenomena emerges when an agent engages in a taboo comparison, but ultimately decides that the risk of detection by her peers is too great to justify choosing the profane outcome, she has still acted outside ISVT’s decision procedure, but just happened to make the *morally-sanctioned* choice. We know

agent makes an immoral choice, they mentally extricate themselves from the community, accepting the guilt and social censure that accompanies the discovery of taboo behavior. Obviously immoral actions can be chosen when the personal costs of a moral obligation are so onerous that an individual sees opting-out of the moral community (publically or incognito) as their best option. However, immoral behavior on one value does not necessarily imply that sacred values closely related to the transgressed value are insincere, nor that past actions in their defense were merely grandstanding. Many Muslims routinely violate sacred values against drinking, but remain genuinely committed to other ideals like charity and piety. Sacred values are isolated from one other, so the addition or subtraction of a sacred value does not necessarily affect one's attitudes towards other sacred objects.

When acting as a member of a group, the agent truly desires the full actualization of her sacred values, but when making a choice as an individual, the subjective costs of sacrifice are sometimes too great for an agent. This is especially true if the taboo act is unseen, since it has no effect on how members conceive of their group, which is where the harm of *public* sacrilege comes from. Of course, if no one sees the violation, no one judges the offender harshly; though if the group finds out about, they feel outrage at the slight—as more than a few Muslim teenagers have found. Within private spheres, agents can represent themselves as *autonomous individuals*, with moral obligations that can be dispensed if the benefit is great enough. The conflicting norms of practical reason that coexist make one's self-representation as a social or individual agent extremely powerful tactic in regulating our conative attitudes.

---

this is the case because, in a slightly different tradeoff, with the profane choice sweetened, an amoral agent would revise his preference ranking.

## 6 CONCLUSION

In this paper, I have argued for introducing a discontinuity between standard actions and sacred actions, using two different practical reasoning systems to predict behavior, described by ISVT and RCT respectively. The separate competencies responsible for action and their rules causes wide variation in the properties of behavior, reflected in the contrast between the sometimes reckless abandon of sacred action and the careful consideration of an important investment or purchase. Similarly, Durkheim argues that people are “homo duplex,” with a nature split into two sides; he writes, “man is double. In him are two beings: an individual being that has its basis in the body and whose sphere of action is strictly limited by this fact, and a social being that represents within us the highest reality in the intellectual and moral realm...” (p. 15). His emphasis on the social origin and purposes of morality usefully frames the distinction between the individualistic profane world, subject to the sciences of economics, and the other collective sacred sphere, which inaugurates laws of a higher caliber. Meeting both our individual and social needs requires striking a balance between the operation of the two competencies, one cannot ruthlessly maximize subjective value at all times, nor can one view every decision as a moral one.

I hope my structure of sacred values affords the reader some insight into these heady times of extremism and polarization, when one might see a “general effervescence that is characteristic of revolutionary or creative epochs” rising to the top of our own societies (Durkheim 1912 p. 213). The toxic play of the sacred and profane that belong to opposed groups becomes apparent in discourse surrounding border walls, symbolic displays of economic sovereignty in the guise of tariffs, and even the separation of immigrant children from their parents. This project can be extended even further by attending to the recent emergences of the Black Lives Matter movement, large-scale anti-capitalist movements, and the increased visibility and popularity of hate groups,

all of which are driven by sacred values tied to racial and gender identities, often times in viscous opposition to other such groups. The relationship between sacred values and both racism and sexism deserve their own, intersecting, analyses, but I could not hope to suitably engage in either for reasons including both the complexities of the issues and my lack of sufficient familiarity with the relevant literatures.

## REFERENCES

- Ainslie, George (2002) "The Effect of Hyperbolic Discounting on Personal Choices," Keynote: APA Conference, 'Personal Choice and Change,' Chicago.
- Atran, Scott (2016). "The Devoted Actor," *Current Anthropology* 57: 13.
- ¶ (2006). "Sacred values, terrorism, and the limits of Rational Choice," *Presented to Department of Homeland Security*
- Baron, Jonathon & Spranca, Mark (1997). "Protected Values," *Organizational Behavior and Human Decision Processes* 70: 1.
- Boot, Martijn (2017). "Problems of Incommensurability," *Social Theory and Practice* 43:2.
- Broome, John (2000). "Incommensurable Values," *Well-Being and Morality: Essays of Honor of James Griffin* [Eds. Crisp and Hooker], Oxford University Press: Oxford.
- Chang, Ruth (1997) "Introduction," in *Incommensurability, Incomparability, and Practical Reason* [Ed. Chang], Harvard University Press: Cambridge.
- (2002) "The Possibility of Parity," *Ethics* 112: 4.
- Chomsky, Noam (1965). *Aspects of the Theory of Syntax*, M.I.T. Press: Cambridge.
- Cohen, L. Jonathan (1981). "Can Human Irrationality be Experimentally Demonstrated?" *Behavioral and Brain Science* 4.
- Davidson, Donald (1963). "Actions, Reasons, and Causes," *The Journal of Philosophy* 60: 23.
- Dehghani, Mostafa (2010). "Sacred values and conflict over Iran's nuclear program," *Judgment and Decision Making* 5.
- De Martino, Benedetto, et. al. (2006). "Frames, Biases, and Rational Decision Making in the Human Brain," *Science* 313: 5787.

- Duc, Corinne, et. al. (2013). "Sacred Values: Trade-off Type Matters," *Journal of Neuroscience, Psychology, and Economics* 6:4.
- Durkheim, Emile (1912). *The Elementary Forms of Religious Life* [Trans. Fields (1995)], Simon & Schuster: New York.
- Finucane, Melissa, et. al. (2000). "The Affect Heuristic in Judgments of Risk and Benefits," *Journal of Behavioral Decision Making* 13.
- Gigerenzer, Gerd & Selten, Richards [Eds.] (2002) *Bounded Rationality: The Adaptive Toolbox*, MIT Press: Cambridge.
- Ginges, Jeremy & Atran, Scott (2009). "Noninstrumental Reasoning over Sacred Values: An Indonesian Case Study," *Psychology of Learning and Motivation* 50.
- Hare, Caspar (2010). "Take the sugar." *Analysis* 70.
- Hasselman, Martin & Tanner, Carmen (2008). "Taboos and conflicts in decision making: Sacred values, decision difficulty, and emotions," *Judgment and Decision Making* 3:1.
- Hofmeyr, Andre, et. al. (2010). "The relationship between addiction and reward bundling: an experiment comparing smokers and non-smokers," *Addiction* 106.
- Hursthouse, Rosalind (1991). "Arational Actions," *The Journal of Philosophy* 88: 2.
- Jeffrey, Richard (1965). *The Logic of Decision*, McGraw-Hill: New York
- Kubrick, Stanley (1987). *Full Metal Jacket* [prod. Kubrick], Warner Bros.: Burbank.
- Nisbett, Richard, et. al. (1983) "The Use of Statistical Heuristics in Everyday Inductive Reasoning," *Psychological Review* 90: 4.
- Peterson, Martin (2009) *An Introduction to Decision Theory*, Cambridge University Press: Cambridge.
- Raz, Joseph (1986). *The Morality of Freedom*, Oxford University Press: Oxford.



- Regenwetter, Michel & Dana, Jason (2011). "Transitivity of Preferences," *Psychological Review* 118: 1.
- Ryan, Erin (2017). "Pro-Life Congressman Adopts Sensible 'Abortion for Mistresses Only' Stance," *The Daily Beast* 10/03/2017.
- Savage, Leonard (1954). *The Foundations of Statistics*, Dover Publications: New York.
- Sheikh, Hammad, et. al. (2012). "Religion, group threat and sacred values," *Judgment and Decision-Making* 7: 12.
- Skryms, Brian (1998). "The Shadow of the Future," *Rational Commitment and Social Justice: Essays for Gregory Kavka* [Ed. Coleman and Morris], Cambridge University Press: New York.
- Smith, Michael (1987) "The Humean Theory of Action," *Mind* 96: 381.
- Sosis, Richard & Alcorta, Candace (2004). "Signalling, Solidarity, and the Sacred: The Evolution of Religious Behavior," *Evolutionary Anthropology* 12: 2.
- Stanford, Kyle (2018). "The difference between ice cream and Nazis: Moral externalization and the evolution of human cooperation," *Behavioral and Brain Sciences* 95.
- Tetlock, Philip, et. al. (2000). "The Psychology of the Unthinkable: Taboo Trade-Offs, Forbidden Base Rates, and Heretical Counterfactuals," *Journal of Personality and Social Psychology* 78: 5.
- (2017). "Sacred versus Pseudo-Sacred Values: How People Cope With Tradeoffs" *American Economic Review: Papers and Proceedings* 107: 5.
- Tversky, Amos (1969). "Intransitivity of Preferences" *Psychological Review* 76:1.
- Tversky, Amos & Kahneman, Daniel (1974). "Judgement under Uncertainty: Heuristics and Biases," *Science* 185: 4157.

- (1979). “Prospect Theory: An Analysis of Decision Under Risk,” *Econometrica* 47: 2.
- (1981). “The Framing of Decisions and the Psychology of Choice,” *Science* 211: 4481.
- (1983). “Extensional Versus Intuitive reasoning: The Conjunction Fallacy in Probability Judgment,” *Psychological Review* 90: 4.

Von Neumann, John & Morgenstern, Oskar (1944). *Theory of Games and Economic Behavior*, Princeton University Press: Princeton.

Wittgenstein, Ludwig (1961). *On Certainty*, Basil-Blackwell: Oxford.

## APPENDIX: A FORMALIZATION OF ISVT

I define a simple logic for decisions involving sacred values using the theoretical resources of ISVT. I introduce two mental attitude predicates, which symbolize the cognitive attitudes of valuing an outcome as sacred or profane. I introduce three relations: first, incomparability ( $>$ ), second, incommensurability ( $\dashv$ ), and third, a change in the attitude type with which an outcome is held ( $\diamond$ ). The final aspect of the theory is definitions of behavior in profane, taboo, and tragic tradeoffs. I derive some implications from these properties that mirror behavior in the world.

### Definition of terms:

$o_n$  : outcome, with 1, 2, 3... specifying identity

$o_{n+}$  : outcome with improved value over  $o_n$

$S_x(o_n)$  :  $o_n$  instantiates a sacred value, with a, b, c... specifying identity

$P(o_n)$  :  $o_n$  instantiates profane value

### Definition of operators:

$X > Y$  : X is preferred to Y

$X = Y$  : X and Y are equally preferable

$X > Y$  : X is sacredly-preferred to Y, where X is a sacred outcome and Y is a profane outcome

$X \dashv Y$  : X and Y are incommensurable, where X and Y are different sacred values

$X \diamond Y$  : X changes attitude-type to Y

$X > Y \rightarrow X$  : X is chosen

### Tradeoffs:

- |   |                   |
|---|-------------------|
| 1. $S_1(o_1) > P(o_2) \rightarrow o_1$  | Taboo tradeoffs   |
| 2. $S_1(o_1) \dashv S_2(o_2) \rightarrow o_1 \text{ or } o_2 \text{ or } \emptyset$ | Tragic tradeoffs  |
| 3. $P(o_1) > P(o_2) \rightarrow o_1$  | Profane tradeoffs |

Applications:

1.  $S_1(o_1) > P(o_2)$

$P(o_3) > P(o_2)$

$\therefore S_1(o_1) > P(o_3) \rightarrow o_1$

2. At  $t_1$ ,  $S_1(o_1) \not\vdash S_2(o_2) \rightarrow o_1$

At  $t_2$ ,  $S_1(o_1) \not\vdash S_2(o_2) \rightarrow o_2$

3. [a]

At  $t_1$ ,  $S_1(o_1) > P(o_2)$

At  $t_2$ ,  $S_1(o_1) \diamond P(o_1)$

$\therefore$  At  $t_3$ ,  $P(o_1) > P(o_2)$ ,  $P(o_2) > P(o_1)$ , or

$P(o_1) = P(o_2)$

Increasing the value of profane outcomes does not result in a taboo action.

Incommensurability implies instability across multiple forced choice situations.

If an agent changes attitude towards an outcome from sacred to profane, nothing is implied about its profane ranking.

[b]

At  $t_1$ ,  $P(o_1) > P(o_2)$

At  $t_2$ ,  $P(o_2) \diamond S_1(o_2)$

$\therefore$  At  $t_3$ ,  $S_1(o_2) > P(o_1) \rightarrow o_2$

If an agent changes attitude towards an outcome from profane to sacred, nothing about its profane ranking affects its new sacred properties.

Small improvements argument:

4. If  $S_1(o_1) = S_2(o_2)$

then  $S_1(o_{1+}) > S_2(o_2)$

$\sim [S_1(o_{1+}) > S_2(o_2)]$

$\therefore S_1(o_1) = S_2(o_2)$  is false

If the two sacred outcomes are equal, then sweetening should result in choice  $o_{1+}$ .

However, no stable preference (uncertainty).

Therefore, the outcomes cannot be equal.

$$5. \quad S_1(o_1) \dashv\vdash S_2(o_2)$$

Implication of small improvement argument.

$$\therefore S_1(o_1) \dashv\vdash S_2(o_2+)$$